HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI
MATEMAATTIS-LUONNONTIETEELLINEN TIEDEKUNTA
MATEMATISK-NATURVETENSKAPLIGA FAKULTETEN
FACULTY OF SCIENCE

Tiedekunta – Fakultet – Faculty Faculty of Science	**Koulutusohjelma – Utbildningsprogram – Degree programme** Master in Particle Physics and Astrophysical Sciences	
Tekijä – Författare – Author Marcos González Ateca		
Työn nimi – Arbetets titel – Title Estimating the two-point correlation function of a 3D galaxy distribution.		
Työn laji – Arbetets art – Level Master thesis	**Aika – Datum – Month and year** February 2020	**Sivumäärä – Sidoantal – Number of pages** 44

Tiivistelmä – Referat – Abstract

The distribution of matter in space is not homogeneous. Large structures such as galaxy groups, clusters or big empty spaces called voids can be observed at large scales in the Universe. The large scale structure of the Universe will depend on both the cosmological parameters and the dynamics of galaxy formation and evolution. One of the main observables that allow us to quantify this structure is the two-point correlation function, with which we can trace different galaxy properties such as luminosity, stellar mass and also, it enables us to track its evolution with redshift.

In galaxy surveys, we do not obtain the location of galaxies in real space. We obtain our data in what it is called redshift space. This redshift space can be defined as a distortion of the real space generated by the redshift introduced by the peculiar velocities of galaxies and from the Hubble expansion of the Universe. Therefore, the distribution of galaxies in redshift space will look different from the one obtained in real space. These differences between both spaces are small but not negligible, and they depend strictly on the cosmology. In this work, we will assume a ΛCDM cosmology.

Therefore, in order to find the different 1-dimensional or 2-dimensional correlations functions, we will use the most updated version of the code provided by the Euclid consortium, which belongs officially to the ESA Euclid mission. Moreover, we will also need different galaxy catalogues. These catalogues have already been simulated and they are called Minerva mocks, which are a set of 300 different cosmological mocks produced with N-body simulations.

Finally, as there is a well-defined relation between real and redshift space, one could also assume that there is a relation between the two-point correlation functions in both real and redshift space. In this project, we will prove that the real-space one-dimensional two-point correlation function, which is the physically meaningful one, can be derived from the two-dimensional two-point correlation function in redshift space following a geometrical procedure independent of approximations. This method, in theory, should work for all distance scales.

Avainsanat – Nyckelord – Keywords
two-point correlation function , redshift space , projected correlation function ,deprojection procedure

Säilytyspaikka – Förvaringställe – Where deposited
E-thesis

Muita tietoja – Övriga uppgifter – Additional information

Master's thesis

Degree Programme in Particle Physics and Astrophysical Sciences
Study track: Elementary particle physics and cosmology

ESTIMATING THE TWO-POINT CORRELATION FUNCTION OF A 3D GALAXY DISTRIBUTION

Marcos González Ateca

07/02/2020

Supervisor : Elina Keihänen

Examiner : Mark Hindmarsh

UNIVERSITY OF HELSINKI
FACULTY OF SCIENCE
P.O.Box 64 (Gustaf Hällströmin katu 2)
00014 University of Helsinki

Contents

1 Abstract

The distribution of matter in space is not homogeneous. Large structures such as galaxy groups, clusters or big empty spaces called voids can be observed at large scales in the Universe. The large scale structure of the Universe will depend on both the cosmological parameters and the dynamics of galaxy formation and evolution. One of the main observables that allows us to quantify the degree of structure is the two-point correlation function, with which we can trace different galaxy properties such as luminosity, stellar mass and also, it enables us to track its evolution with redshift.

Nowadays, the large scale structure of the Universe is thought to come from small fluctuations, in the temperature of the cosmic microwave background (CMB), that have grown into the observable structure of today through gravitational instability. Therefore, measuring these structures could help us to constrain both, the true cosmology of our Universe and the galaxy dynamics and evolution.

As mentioned before, using the two-point correlation function, the evolution of the large scale structure with redshift can be estimated. In galaxy surveys, we do not obtain the location of galaxies in real space. We obtain our data in what it is called *redshift space*. This redshift space can be defined as a distortion of the real space generated by the redshift introduced by the peculiar velocities of galaxies and from the Hubble expansion of the Universe. Therefore, the distribution of galaxies in redshift space will look different from the one obtained in real space. These differences between both spaces are small but not negligible, and they depend strictly on the cosmology. In this work, we will assume a ΛCDM cosmology.

Therefore, in order to find the different 1-dimensional or 2-dimensional correlations functions, the most updated version of the code provided by the Euclid consortium, which belongs officially to the ESA Euclid mission, will be used. Moreover, we will also need different galaxy catalogues. These catalogues have already been simulated and they are called Minerva mocks, which are a set of 300 different cosmological mocks produced with N-body simulations. These mocks will have a cubic shape and periodic boundary conditions will be assumed. During this project, the most updated values of the cosmological parameters have been introduced in the code so that the results are as accurate as possible.

Finally, as there is a well-defined relation between real and redshift space, one could also assume that there is a relation between the two-point correlation functions in both real and redshift space. In this project, we will try to prove that the real-space one-dimensional two-point correlation function can be derived from the two-dimensional two-point correlation function in redshift space following a geometrical procedure independent of approximations. This method, in theory, should work for all scales.

2 Objectives

GENERAL OBJECTIVE

- Prove that the one-dimensional two-point correlation function in real space can be derived from redshift-space anisotropic two-point correlation function.

SPECIFIC OBJECTIVES

- Compute, using the Euclid code, the one-dimensional and two-dimensional correlation functions in both redshift and real space.
- Compute the projected correlation functions obtained from real and redshift space correlation functions and compare them.
- Deproject the projected correlation functions to obtain the one-dimensional correlation function in real space.
- Find a proper value for the maximum integration limit π_{max} that has to be used to calculate projected correlation functions.
- Briefly discuss the implications of these calculations for future observations.

3 Our Universe

3.1 Friedmann equations

We consider now a homogeneous and isotropic Universe. This model is usually called the Friedmann-Lemaître-Robertson-Walker or Friedmann-Robertson-Walker (FRW) Universe. There are scientific proofs that indeed, the Universe is rather homogeneous for large scales of about 100Mpc or more. This means that for those scales, the Universe practically looks the same no matter the direction you look.

The metric of this kind of Universe in terms of the comoving coordinates is given by,

$$ds^2 = -c^2dt^2 + a(t)^2\left[dr^2 + r^2(d\theta^2 + \sin\theta^2 d\phi^2)\right] \ , \tag{1}$$

where c is the speed of light, from now on we will consider $c = 1$, and $a(t)$ is the so-called *scale factor*, which is usually normalized to have a value of 1 at present time, $a(t_0) = 1$.

This kind of Universe is no static, it is an expanding Universe. Applying Einstein equation, which describes the curvature of spacetime in terms of matter and energy, to this kind of Universe, one gets the Friedmann equations,

$$\left(\frac{\dot{a}}{a}\right)^2 + K/a^2 = \frac{8\pi G}{3}\rho \tag{2}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) \ , \tag{3}$$

with G being Newton's gravitational constant, K the curvature of the Universe, ρ the total energy density and p the total pressure of the fluid. If we now apply the assumption made at the beginning, which states that the Universe is homogeneous, we deduct that the pressure and the density of matter, or energy, must be the same everywhere. Therefore, there is only a time dependence for these quantities and thus, if we use the Hubble parameter $H = \dot{a}/a$, we can rewrite Friedmann equations to be,

$$H^2 = \frac{8\pi G}{3}\rho - \frac{K}{a^2} \tag{4}$$

$$\dot{H} = -4\pi G(\rho + p) + \frac{K}{a^2} \ . \tag{5}$$

From the equivalence principle of general relativity, it is required that locally, both energy and momentum are conserved. From this statement, the energy continuity equation can be derived from equations 4 and 5 ,

$$\dot{\rho} = -3(\rho + p)\frac{\dot{a}}{a} \ . \tag{6}$$

In order to solve Friedmann equations, we will need an equation that relates the pressure p with the density ρ. This equation is called the *equation of state* and it can be written as,

$$p = w\rho \ . \tag{7}$$

If we now assume that the Universe contains non-relativistic matter and radiation, mostly in the cosmic microwave background, we can generalize our equation of state to take all the different energy components into account so that,

$$\rho = \sum_i \rho_i \qquad and \qquad p_i = w_i \rho_i \ , \tag{8}$$

and thus if for simplicity, we assume a constant w, we can integrate the energy continuity equation 6 to obtain,

$$\rho_i \propto a^{-3(1+w_i)} \ . \tag{9}$$

The result above is basically correct for all the stages of the Universe after the Big Bang Nucleosynthesis, where we have,

$$\rho = \rho_r + \rho_m + \rho_{vac} \qquad with \qquad w_r = 1/3, w_m = 0, w_{vac} = -1 \ . \tag{10}$$

We can now define the cosmological parameters by dividing each density component by the critical density, defined as $\rho_{crit} = \frac{3H_0^2}{8\pi G}$. Therefore,

$$\Omega = \Omega_r(t) + \Omega_m(t) + \Omega_\Lambda(t) + \Omega_k(t) = 1 \tag{11}$$

and then, equation 4 can be easily rewritten as,

$$H(a) = H_0\sqrt{\Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda} \ , \tag{12}$$

where $H_0 = 70\frac{km}{sMpc}$ is the so-called Hubble constant.

3.2 Redshift

Redshift is defined as the effect for which the wavelength of a photon is stretched, so the light is seen as shifted towards the part of the spectrum with longer wavelengths. The magnitude of this change is the difference between the observed and emitted wavelength divided by the emitted wavelength,

$$z = \frac{\lambda_o - \lambda_e}{\lambda_e} \qquad \Rightarrow \qquad 1 + z = \frac{\lambda_o}{\lambda_e} \ , \tag{13}$$

where the o subscript means observed and the e one means emitted. In our kind of Universe, FRW Universe, redshift is caused by the expansion of the Universe and it can be related to the scale factor quite straightforward,

$$1 + z = \frac{a_o}{a_e} \qquad \Rightarrow \qquad a(z) = \frac{1}{1+z} \ , \tag{14}$$

with $a_o = a(t_o) = 1$ and $a_e = a(t) = a$. Using the previous relation, equation 12 can be rewritten in terms of the redshift,

$$H(z) = H_0 \sqrt{\Omega_r (1+z)^4 + \Omega_m (1+z)^3 + \Omega_k (1+z)^2 + \Omega_\Lambda} \ . \tag{15}$$

In all our calculations, we have made use of the most updated values of the cosmological and Hubble parameters. These values are the ones used in the Euclid Consortium code and have values of,

$$\begin{aligned} \Omega_m &= 0.278 \\ \Omega_\Lambda &= 0.763 \\ \Omega_r = \Omega_k &= 0.0 \ . \end{aligned} \tag{16}$$

Now that we know how does the Hubble parameter evolve with redshift, the relation between distance and redshift can be calculated. Therefore, if we consider a photon moving radially towards an observer, we will have that the distance-redshift relation using the above cosmological parameters gives,

$$r(z) = H_0^{-1} \int_0^z \frac{dz'}{\sqrt{\Omega_m (1+z')^3 + \Omega_\Lambda}} \ , \tag{17}$$

where r is the comoving distance. For very small redshifts, $z \ll 1$, we could make use of the linear Hubble relation, $z = H_0 r$. This approximation for small redshifts ignores peculiar velocities. In 2022, the Euclid satellite will begin collecting data. The galaxy catalogues that will be registered will have data of galaxies up to redshifts close to 2, and thus, the approximation of

the linear Hubble relation will not work. We have to add the effect of the peculiar velocities. Therefore, for an individual galaxy,

$$z = H_0 r + v_r \qquad with \qquad v_r = \hat{\mathbf{r}} \cdot \mathbf{v} \ , \tag{18}$$

where, again, we also take into account the contribution on the radial direction. If we would want to measure these peculiar velocities, we would need to differentiate between cosmological and observed redshifts in our surveys. The cosmological redshifts, z_{cos}, are due solely to the expansion of the Universe, while the observed redshift, z_{obs}, is a contribution of both the cosmological one and the effect of peculiar radial velocities. Therefore, from reference [33], we obtain an estimation of the non-relativistic peculiar velocities, $v_r \ll 1$, which is,

$$v_r = \frac{z_{obs} - z_{cos}}{1 + z_{cos}} \ . \tag{19}$$

Finally, using $H_0^{-1} = 1$, we can define the redshift vector $\mathbf{z}$ for each individual galaxy,

$$\mathbf{z} \equiv z\hat{\mathbf{r}} = \mathbf{r} + v_r\hat{\mathbf{r}} = \left(1 + \frac{v_r}{r}\right)\mathbf{r} \ . \tag{20}$$

This redshift vector gives the position of the galaxy in *redshift space*. As it was explained in the abstract, when a sky survey is conducted, the position obtained for galaxies is not in real space, but redshift space. The differences between these two are quite small, but they have to be taken into account. These differences are produced, mainly, by the Hubble expansion of the Universe and the peculiar velocities of galaxies. From equation 20, it can easily be seen that the contribution of the peculiar radial velocities will make the distribution of galaxies in redshift space look different from the one in real space.

Till now, we have treated the Universe as isotropic and homogeneous. Now, we will move on to study small inhomogeneities in the beginning of our Universe which led to the big inhomogeneities in the Universe that can be observed today.

3.3 Inhomogeneus Universe

The most accepted idea related to the origin of structure in the Universe, states that during the inflationary period of the Universe, initial density perturbations were produced by quantum fluctuations of the inflaton field.

Then, these fluctuations of the density field have been evolving until today following the laws of deterministic physics. When we now look to our Local Universe, large inhomogeneities can easily be seen, from galaxy clusters or groups to voids. If we go deeper into smaller scales, these inhomogeneities only grow more prominent, having different galactic components, globular clusters, interstellar medium, stars or, for example, planets.

In October 2013, the ESA's Planck satellite ended its period of data collection of the cosmic microwave background. From that mission, one of the main conclusions was that the primordial Universe was much more homogeneous than the one we see today. The magnitude of the inhomogeneities was close to $\sim 10^{-5}$ at the time of recombination. Now, we will give a summary of how these inhomogeneities grew from that small variation on the density field to the large scale structure we can observe today.

3.3.1 Inhomogeneity

We are now going to explain the evolution of the density field $\rho(\mathbf{x})$ in time or, in another words, the theory behind structure formation. From this point on, we will consider this density field to be ergodic and statistically homogeneous and isotropic. We define the *density perturbation field* as,

$$\delta(\mathbf{x}) \equiv \frac{\rho(\mathbf{x}) - \langle\rho\rangle}{\langle\rho\rangle} \ , \tag{21}$$

where $\langle\rho\rangle$ is the average density field over all space. From statistical homogeneity, $\langle\rho(\mathbf{x})\rangle = \langle\rho\rangle$, and therefore, $\langle\delta\rangle = 0$, which means that $\langle\delta\rangle$ can not be used as a measure of the inhomogeneity.

In order to describe the evolution of the density field, some assumptions have to be made so that we can explain it using classical physics. These assumptions are: (i) The matter density field can not be relativistic, (ii) the peculiar velocities are also non-relativistic, and finally, (iii) the considered distance scales are much smaller than the Hubble length. If all these assumptions are fulfilled, we say that we are in the *Newtonian approximation* and thus, linear perturbation theory can be applied. Therefore, using linear perturbation theory, the equations we obtain for a perfect fluid in terms of $\delta(\mathbf{x})$ are,

$$\begin{aligned} \dot{\delta} + \frac{\nabla \cdot \mathbf{v}}{a} &= 0 \\ \dot{\mathbf{v}} + Hv + \frac{\nabla P}{\rho a} + \frac{\nabla \Phi}{a} &= 0 \\ \nabla^2 \Phi &= 4\pi G a^2 \bar{\rho} \delta \ , \end{aligned} \tag{22}$$

where Φ is the gravitational potential, a the scale factor, G is the gravitational constant and P is the pressure on the fluid. The first equation is called the *continuity equation*, the second one the *Euler equation* and the last one, the *Poisson equation*. Now, if we differentiate the continuity equation and introduce the Euler and Poisson equations, we reach a second-order differential equation, usually called *master equation*, that describes the evolution of density perturbations. Due to the approximations made, this equation will only work in the linear regime, ($\delta \ll 1$), for a non-relativistic fluid,

$$\ddot{\delta} + 2H\dot{\delta} - 4\pi G a^2 \bar{\rho} \delta = 0 \ , \tag{23}$$

where we have neglected the entropy perturbations and assumed a pressureless fluid. From the second term on this equation, called *Hubble drag* term, we can see how does the expansion of the Universe suppress the growth of the perturbation. In the same way, the last term called *gravitational term*, expresses how does gravity promote the expansion of these density perturbations.

From equation 23 it can be observed that, as the equation is a second-order differential equation, there are 2 different solutions, so that the general solution will be a linear combination of these solutions,

$$\delta(\mathbf{x}, t) = \delta_1(\mathbf{x}, t) + \delta_2(\mathbf{x}, t) \tag{24}$$

Moreover, in equation 23, all the derivatives are with respect to time and thus, there is no $\mathbf{x}$ dependence. Therefore, these solutions can be separated into two functions, one with time dependence, D(t), and another with spatial dependence $\Delta(\mathbf{x})$, so that, $\delta(\mathbf{x}, t) = D(t)\Delta(\mathbf{x})$. Therefore, we can write a general solution to the equation 23 as,

$$\delta(\mathbf{x}, t) = D_1(t)\Delta_1(\mathbf{x}) + D_2(t)\Delta_2(\mathbf{x}) \ . \tag{25}$$

In this equation, the $\Delta(\mathbf{x})$ functions are the functions that define the primordial density distribution, and the D(t) functions are the so-called

density growth factors. If we now consider an initial density distribution in the matter era, we can conclude that the topology of the matter distribution will not change as the perturbations evolve, and also, that the growth rate of the density perturbations will be equal at every location. For this kind of distribution, we will have two independent growth factors, whose values are,

$$D_1 \propto a(t) \tag{26}$$

$$D_2 \propto a(t)^{-3/2} \ , \tag{27}$$

which are usually called growing and decaying modes. Finally, it has to be remarked that in the linear perturbation theory, the overdense regions, regions with $\delta > 0$, will grow and, on the contrary, underdense regions will become more underdense. After some time, δ will grow to be bigger than 1 and from that moment onwards, nonlinear perturbation theory will have to be applied. For these cases, higher-order perturbation theory could be applied as well as the Zeldovich approximation, where a homogeneous and collisionless initial matter distribution is assumed. However, in the end, we will need some kind of numerical simulations to solve the equations due to the impossibility of doing it analytically.

Now that we know how these inhomogeneities $\delta(\mathbf{x}, t)$ evolve in real space, it is useful to expand them as a Fourier series as,

$$\delta(\mathbf{x}) = \sum_{\mathbf{k}} \delta_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} \ , \tag{28}$$

where $\delta_{\text{-}\mathbf{k}} = \delta^*_{\mathbf{k}}$, k is the wavenumber, which is related to the wavelength of the fluctuation by $k = \frac{2\pi}{\lambda}$, and,

$$\delta_{\mathbf{k}} = \frac{1}{V} \int_V \delta(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} d^3x \ . \tag{29}$$

It can easily be seen that, as

$$\langle \delta(\mathbf{x}) \rangle = 0 \Rightarrow \langle \delta_{\mathbf{k}} \rangle = 0 \ . \tag{30}$$

From here, it is explained that $\langle \delta \rangle$ can not be used if we want to measure the magnitude of the inhomogeneity. Therefore, another quantity must be found. The easiest choice for this quantity is the square of the perturbation. This term would be positive everywhere and thus, its average could not be

0 as in the previous example. This term will be called *density 2-point autocorrelation function*, or simply, *two-point correlation function* and it will be a critical factor to measure the structure formation and evolution in the Universe.

3.3.2 Two-point correlation function

The two-point correlation function or, 2PCF, is one of the most important statistical tools that help us quantifying the excess probability of finding one object within a specific distance to another object against that of a random and unclustered distribution. The two-point correlation function, in general, traces the amplitude of clustering as a function of scale. This 2PCF can be used, irrespective of whether we are measuring the clustering of stars, galaxies or clusters of galaxies.

The 2PCF is defined as,

$$\xi(\mathbf{x}_1, \mathbf{x}_2) = \langle \delta(\mathbf{x}_1)\delta(\mathbf{x}_2) \rangle \ . \tag{31}$$

This quantity will be positive if the density perturbation has an identical sign in both $\mathbf{x}_1$ and $\mathbf{x}_2$ whereas it will be negative if we have different signs in those locations, which means that there is an under density in one place and an overdensity in the other. Furthermore, it can be proven the fact that density perturbations at different locations are intercorrelated between each other. Applying statistical homogeneity, $\mathbf{r} \equiv \mathbf{x}_2 - \mathbf{x}_1$, we can rewrite the two-point correlation function as a more commonly seen expression,

$$\xi(\mathbf{r}) = \langle \delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r}) \rangle \ . \tag{32}$$

From statistical isotropy, $\xi(\mathbf{r}) = \xi(r)$ can be inferred. As it has already been said, the correlation function represents the excess probability of finding points with a determined separation r. If we would only consider a homogeneous Poisson process, we would have that the probability of finding two points in different volumes would be,

$$dP_{12} = n^2 dV_1 dV_2 \ , \tag{33}$$

where n is the mean object number density. However, if we are considering an inhomogeneous process, in the case of clustering there will be a difference

with respect to the Poisson distribution due to the underlying matter density perturbations. Therefore, it will be given by,

$$dP_{12} = n^2[1 + \xi(r)]dV_1 dV_2 \ , \tag{34}$$

and thus, we conclude that $\xi(r)$, measures the level of clustering, its excess or defect, compared to a homogeneous distribution in space. Then, if there is clustering in space at a certain distance r, we will have that $\xi(r) > 0$. If there is no visible structure so that we would have a homogeneous distribution, $\xi(r) = 0$ and, if points tend to avoid each other, $\xi(r) < 0$.

We can now define the correlation function for a single realization as a volume average. This is what we will usually end up calculating. It is defined as,

$$\hat{\xi}(\mathbf{r}) = \frac{1}{V}\int d^3x \ \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r}) \ . \tag{35}$$

If we now integrate this expression over all space assuming periodic boundary conditions, which are a set of boundary conditions chosen to approximate a very large, or infinite, system by using a small part of it called *unit cell*, we get what it is usually called *integral constraint*,

$$\int d^3r \ \hat{\xi}(\mathbf{r}) = \frac{1}{V}\int d^3x \ \delta(\mathbf{x}) \int d^3r \ \delta(\mathbf{x}+\mathbf{r}) = 0 \ . \tag{36}$$

Therefore, from this integral constraint, we conclude that the correlation function $\xi(r)$, must become negative somewhere so that at a certain distance from an overdense region it is more likely to find an underdense region. If we keep increasing the distance, we will have that the correlation function oscillates around zero while becoming smaller in amplitude. However, the most scientifically interesting part of the two-point correlation functions is confined to the small values of r, where the correlation function is positive.

In Fourier space, the correlation $\langle \delta^*_{\mathbf{k}} \delta_{\mathbf{k}'} \rangle$ can also be found. It is important to emphasise that the quantities in Fourier space preserve the same information as the ones in real space but, sometimes, they are easier to deal with .Using equation 29, the result is,

$$\begin{aligned}
\langle \delta^*_{\mathbf{k}} \delta_{\mathbf{k}'} \rangle &= \frac{1}{V^2} \int d^3x e^{i\mathbf{k}\cdot\mathbf{x}} \int d^3x' e^{-i\mathbf{k}'\cdot\mathbf{x}'} \langle \delta(\mathbf{x}) \delta(\mathbf{x}') \rangle \\
&= \frac{1}{V^2} \int d^3x e^{i\mathbf{k}\cdot\mathbf{x}} \int d^3r e^{-i\mathbf{k}'\cdot(\mathbf{x}+\mathbf{r})} \langle \delta(\mathbf{x}) \delta(\mathbf{x}+\mathbf{r}) \rangle \\
&= \frac{1}{V^2} \int d^3r e^{-i\mathbf{k}'\cdot\mathbf{r}} \xi(\mathbf{r}) \int d^3x e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}} \\
&= \frac{1}{V} \delta_{\mathbf{k}\mathbf{k}'} \int d^3r e^{-i\mathbf{k}\cdot\mathbf{r}} \xi(\mathbf{r}) = \frac{1}{V} \delta_{\ \mathbf{k}\mathbf{k}'} P(\mathbf{k}) \ . \qquad (37)
\end{aligned}$$

From statistical homogeneity we deduce that the Fourier coefficients $\delta_{\mathbf{k}}$ are uncorrelated. Moreover, the *power spectrum* quantity has been introduced in the previous equation. The power spectrum of $\delta(\mathbf{x})$ gives the variance of $\delta_{\mathbf{k}}$,

$$P(\mathbf{k}) \equiv V \langle |\delta_{\mathbf{k}}|^2 \rangle = \int d^3r e^{-i\mathbf{k}\cdot\mathbf{r}} \xi(\mathbf{r}) \ . \qquad (38)$$

One important feature of the power spectrum is that it is always positive, and that is one of the main reasons why, when dealing with large scale perturbations, it is more used than correlation functions. Again, as we did with the correlation function, if we apply statistical isotropy, it can be proven that $P(\mathbf{k}) = P(k)$. Moreover, working with the power spectrum instead of working with correlation functions is also helpful because the power spectrum is the quantity predicted by the theories of inflation and structure formation. However, the correlation function is closer to the observations than the power spectrum. Therefore, for small distance scales, density perturbations are more commonly discussed in terms of $\xi(r)$ and for large distance scales, where $\xi(r)$ is quite small, the power spectrum P(k) is more commonly used.

Another interesting property of the two-point correlation functions is that for certain distance ranges, it can be approximated by a power-law shape function. Based on [6], for distances smaller than $10h^{-1}$ Mpc, we would have that the correlation function is well approximated by,

$$\xi(r) = \left(\frac{r_0}{r}\right)^{\gamma} \ . \qquad (39)$$

In the previous reference, the estimate value of gamma is $\gamma = 1.74 \pm 0.04$ and, the point r_0 is the point where the value of the correlation function equals unity, which corresponds to $r_0 = 5.4 \pm 0.3 h^{-1}$Mpc. A value of $h = 0.70$

is considered for the Hubble parameter.

In the same way, the power spectrum can also be approximated with a power-law function. In this case, it would be given by,

$$P(k) = A^2 \left(\frac{k}{k_p}\right)^n \ , \tag{40}$$

where k_p is the so-called *pivot scale* and $A = \sqrt{P(k_p)}$. Another quantity is introduced, it is usually called spectral index, n(k), and it is defined as,

$$n(k) = \frac{d \ln \mathrm{P}}{d \ln \mathrm{k}} \ , \tag{41}$$

where in the power-law case for the power spectrum gives a constant value of the spectral index $n(k) = const = n$.

Another point that has to be mentioned, is the fact that the two-point correlation function will be affected by the peculiar velocities of the galaxies and, therefore, this correlation function will be different in real or redshift space. In practice, at the time of measuring this function, we will separate the effects and distortions induced by redshift from the spatial correlations by assuming a two-dimensional correlation function, $\xi(r_p, r_\pi)$, where r_p is the separation between each galaxy pair in the direction perpendicular to the line-of-sight, and r_π the parallel to the line of sight direction. Following reference [8], these two separations can be defined as,

$$r_\pi = \frac{\mathbf{s} \cdot \mathbf{l}}{|l|} \tag{42}$$

$$r_p^2 = \mathbf{s} \cdot \mathbf{s} - r_\pi^2 \ , \tag{43}$$

where, $\mathbf{l} = \frac{1}{2}(\mathbf{s}_1 + \mathbf{s}_2)$, $\mathbf{s} = \mathbf{s}_1 - \mathbf{s}_2$ and $\mathbf{s}_1$, $\mathbf{s}_2$ are the positions of the selected pair of galaxies in redshift space. All these definitions can be seen graphically in figure 1.

Another important approximation that will be made is the so-called *plane-parallel approximation*. In this case, we assume the parallel and perpendicular distances between the galaxies, r_π and r_p, to be much smaller than the distance between the observer and the galaxies in the line-of-sight direction, $r_p, r_\pi \ll l$. In other words, galaxies are assumed to be far enough

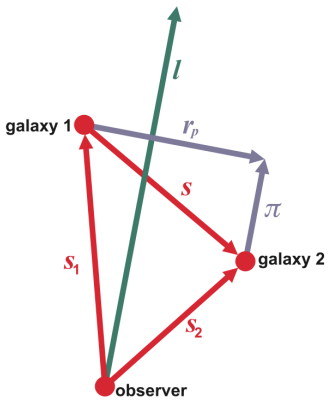

Figure 1: Illustration of the different components of the separation between two galaxies in redshift space.

so that the displacements induced by peculiar velocities are, indeed, parallel. When using this plane-parallel approximation in a galaxy survey, we will be restricted to consider galaxies separated at most by some angle on the sky. As it is stated in reference [23], almost every survey or study related with large-scale redshift space distortions has assumed this approximation.

Therefore, if we calculate the two-dimensional 2PCF, also called anisotropic 2PCF, in real space, we will have to see circular contours that should only depend on the scalar distance between each galaxy pair. On the other hand, if we would calculate it in redshift space, we would see the effects of peculiar velocities of galaxies imprinted in the correlation function. There will be two main effects or distortions caused by these peculiar velocities, at small scales, we will have that the correlation contours will be elongated along the line of sight direction, while at big scales, we will see a compression in the r_π direction. This effect will be proven later on when we start with the two-point correlation function calculations.

Now, that the theory behind correlation functions has been explained, we will move on to explain the most updated methods that are being used nowadays to determine these functions.

3.3.3 Measuring the Two-point correlation function

As we explained previously, the 2PCF is a statistical tool defined as the excess probability of finding two different objects at a certain distance r located at volume elements dV_1 and dV_2, with respect to a Poisson random distribution. From different bibliography references, we can see that different estimators have been proposed throughout the years. These estimators require the simulation of a random data catalogue, with which we can compare the real data catalogue that we have obtained from observations.

Once we have both catalogues, what we have to do is finding the number of pairs of galaxies, DD(r) or RR(r) depending on whether we are on the data catalogue or in the random one, with a separation $|\mathbf{r}_2 - \mathbf{r}_1| \in r \pm \frac{1}{2}\Delta r$, where Δr is called the bin width. Furthermore, there is an unavoidable randomness, in both catalogues, that introduces a random error in our estimators. Therefore, we will try to maximize the number of galaxy points in our catalogues to reduce this error. However, this is not always easy or cheap to do in the data catalogue and, therefore, normally we will use a random catalogue with many more points that the data catalogue, $N_R \gg N_D$.

Now, we recover the most important estimators used to find the two-point correlation functions. These estimators are defined as,

$$1 + \hat{\xi}_{DP}(r) = \frac{2N_R}{N_D - 1}\frac{DD(r)}{DR(r)} \tag{44}$$

$$1 + \hat{\xi}_{H}(r) = \frac{4N_D N_R}{(N_D - 1)(N_R - 1)}\frac{DD(r)RR(r)}{DR(r)^2} \tag{45}$$

$$\hat{\xi}_{LS}(r) = \frac{N_R(N_R - 1)}{N_D(N_D - 1)}\frac{DD(r)}{RR(r)} - \frac{N_R - 1}{N_D}\frac{DR(r)}{RR(r)} + 1\ , \tag{46}$$

where a new quantity DR(r) has been introduced. This quantity called, *data-random pairs*, helps us deal with the edge effects induced by the fact that the galaxy survey only covers a finite volume of space in a single realization. These three estimators have been taken from their original papers, which are the references [4], [6] and [7] of the bibliography. $\hat{\xi}_{DP}$ refers to the Davis-Peebles estimator, $\hat{\xi}_H$ to the Hamilton one and, finally, $\hat{\xi}_{LS}$ to the Landy&Szalay estimator.

These estimators do not give a perfectly precise value of the correlation function. Indeed, some of them introduce a bias in the estimate. Further-

more, the variance of these estimators will have to be found too. Following the paper [4], published by Stephen D. Landy and Alexander S. Szalay back in 1993, we can calculate these quantities for the different estimators. It is important to notice that in that paper, Landy&Szalay apply their theory to calculate the bias and variance of the angular galaxy correlation function estimator. However, the process will be the same for our correlation functions.

As it is done in the paper, we can express the three different pair counts DD(r), DR(r) and RR(r), in terms of the fluctuations around its means as,

$$\begin{aligned} DD(r) &= \langle DD(r)\rangle[1+\alpha(r)] && (47)\\ DR(r) &= \langle DR(r)\rangle[1+\beta(r)] && (48)\\ RR(r) &= \langle RR(r)\rangle[1+\gamma(r)]\ , && (49) \end{aligned}$$

where we assume that the fluctuations are small and thus, $\langle\alpha\rangle = \langle\beta\rangle = \langle\gamma\rangle = 0$. Furthermore, as it is assumed that the number of random points N_R is much larger than the number of data points N_D, we can also approximate $\gamma \approx 0$. Also, if a high number of galaxy pairs in each distance bin is assumed, the fluctuation around the mean of DD(r) and DR(r) will be small and, consequently, α and β will be small and will only be taken into account up to second order. From the Landy&Szalay paper, we have that,

$$\begin{aligned} \langle\alpha^2\rangle &= \frac{\langle DD\cdot DD\rangle - \langle DD\rangle^2}{\langle DD\rangle^2} && (50)\\ \langle\beta^2\rangle &= \frac{\langle DR\cdot DR\rangle - \langle DR\rangle^2}{\langle DR^2\rangle} && (51)\\ \langle\alpha\beta\rangle &= \frac{\langle DR\cdot DR\rangle - \langle DD\rangle\langle DR\rangle}{\langle DD\rangle\langle DR\rangle}\ , && (52) \end{aligned}$$

and the expectation values for the different pair counts DD(r), DR(r) and RR(r) are,

$$\begin{aligned} \langle DD(r)\rangle &= \frac{N_d(N_d-1)}{2}G_p(r)[1+\xi(r)] && (53)\\ \langle DR(r)\rangle &= N_dN_rG_p(r) && (54)\\ \langle RR(r)\rangle &= \frac{N_r(N_r-1)}{2}G_p(r)\ , && (55) \end{aligned}$$

where $G_p(r)$ is the fraction of microcell pairs with separation r defined as,

$$G_p(r) = \frac{2}{K(K-1)} \sum_{i<j}^{K} \Theta_{ij}(r) \ , \tag{56}$$

with K the number of microcells. We will also need to know the second moment of the pair counts, from where we will obtain the results for equations 50, 51 and 52. From Landy&Szalay,

$$\langle DD \cdot DD \rangle \approx \frac{N_D(N_D-1)(N_D-2)(N_D-3)}{4} G_p(r)^2 \tag{57}$$

$$+N_D(N_D-1)(N_D-2)G_t(r) + \frac{N_D(N_D-1)}{2} G_p(r) \tag{58}$$

where very weak correlations have been assumed and,

$$\langle DR \cdot DR \rangle = N_D N_R^2 [G_p(r)^2 (N_D-1) + G_t(r)] \tag{59}$$

$$\langle DD \cdot DR \rangle = \frac{N_D(N_D-1)}{2} N_R [(N_D-2)G_p(r)^2 + 2G_t(r)] \ , \tag{60}$$

being $G_t(r)$ the number of all triplets of microcells, with $\Theta_{ik}(r)\Theta_{jk}(r)$ the factor that selects those terms where the other two points are separated by a distance r from the shared point, defined as

$$G_t(r) = \frac{1}{K(K-1)(K-2)} \sum_{ijk} \Theta_{ik}(r)\Theta_{jk}(r) \ . \tag{61}$$

Now, we have everything we need to rewrite the values we had for the expectation values of the second-order fluctuations. The final result is,

$$\langle \alpha^2 \rangle \approx \frac{2}{N_D(N_D-1)} \left\{ 2(N_D-2) \left[\frac{G_t(r)}{G_p(r)^2} - 1 \right] + \frac{1}{G_p(r)} - 1 \right\} \tag{62}$$

$$\langle \beta^2 \rangle = \frac{1}{N_D} \left[\frac{G_t(r)}{G_p(r)^2} - 1 \right] \tag{63}$$

$$\langle \alpha\beta \rangle = \frac{2}{N_D} \left[\frac{G_t(r)}{G_p(r)^2} - 1 \right] \ . \tag{64}$$

As it is done in the paper, two new variables d and t can be introduced to make these formulas easier. They are defined as,

$$p = \frac{2}{N_D(N_d-1)}\left[\frac{1}{G_p(r)} - 2\frac{G_t(r)}{G_p(r)^2} + 1\right] \approx \frac{2}{N_D(N_D-1)G_p(r)} \quad (65)$$

$$t = \frac{1}{N_D}\left[\frac{G_t(r)}{G_p(r)^2} - 1\right] \quad (66)$$

where $G_t(r) \simeq G_p(r)^2$ is assumed. Therefore,

$$\langle\alpha^2\rangle = 4t + p \quad (67)$$

$$\langle\beta^2\rangle = t \quad (68)$$

$$\langle\alpha\beta\rangle = 2t\ . \quad (69)$$

We now move on to define the normalized pair counts, which will be used to rewrite the three different estimators of the two-point correlation function defined in 44, 45 and 46. These normalized pair counts are defined as,

$$d(r) = \frac{DD(r)}{\frac{1}{2}N_D(N_D-1)G_p(r)} \Rightarrow \langle d\rangle = \frac{1+\xi(r)}{1+\xi_V} \quad (70)$$

$$x(r) = \frac{DR(r)}{N_D N_R G_p(r)} \Rightarrow \langle x\rangle = 1\ , \quad (71)$$

where ξ_V is the *mean expected correlation function over the survey volume*,

$$\xi_V = \sum_r G_p(r)\xi(r)\ . \quad (72)$$

Making use of these definitions, the estimators can be written in a more simple and easy way. The result is,

$$1 + \hat{\xi}_{DP}(r) = \frac{d(r)}{x(r)} \quad (73)$$

$$1 + \hat{\xi}_{H}(r) = \frac{d(r)}{x(r)^2} \quad (74)$$

$$1 + \hat{\xi}_{LS}(r) = d(r) - 2x(r) + 2\ . \quad (75)$$

We now are interested in finding the bias and variance of these estimators. We will solve it for the Landy&Szalay estimator, which is the one that

the Euclid code uses to quantify the two-point correlation function. The bias of this estimator can be very easily calculated by,

$$\begin{aligned} \langle 1+\hat{\xi}_{LS}(r)\rangle &= 1+\langle\hat{\xi}_{LS}(r)\rangle = \langle d(r)-2x(r)+2\rangle \\ &= \langle d(r)\rangle - 2\langle x(r)\rangle + 2 = \langle d(r)\rangle = \langle d\rangle \ . \end{aligned} \tag{76}$$

Therefore, we conclude that the Landy&Szalay estimator is an unbiased estimator except for the $1+\xi_V$ term that appears in equation 70. Now, we want to find the variance of this estimator. Making use of,

$$d = \langle d\rangle(1+\alpha) \qquad \text{and} \qquad x = \langle x\rangle(1+\beta) \ , \tag{77}$$

our estimator can be rewritten as,

$$\begin{aligned} 1+\hat{\xi}_{LS}(r) &= d(r)-2x(r)+2 = \langle d\rangle(1+\alpha)-2\langle x\rangle(1+\beta)+2 \\ &= \langle d\rangle(1+\alpha)-2\beta \ . \end{aligned} \tag{78}$$

In order to find the variance we previously need to find,

$$\langle(1+\hat{\xi}_{LS})^2\rangle = \langle\langle d\rangle^2(\alpha^2+2\alpha+1)-4\langle d\rangle(\beta+\alpha\beta)+4\beta^2\rangle \tag{79}$$

$$= \langle d\rangle^2(1+\langle\alpha^2\rangle)-4\langle d\rangle\langle\alpha\beta\rangle+4\langle\beta^2\rangle \tag{80}$$

$$= \langle d\rangle^2(1+4t+p)-8\langle d\rangle t+4t \ , \tag{81}$$

where the fact that the fluctuations are small $\langle\alpha\rangle = \langle\beta\rangle = 0$ and equations 67, 68 and 69 have been used. Finally, the variance of the estimator is then given by,

$$\begin{aligned} Var(\hat{\xi}_{LS}) &= \langle(1+\hat{\xi}_{LS})^2\rangle - \langle(1+\hat{\xi}_{LS})\rangle^2 \\ &= \langle d\rangle^2 + \langle d\rangle^2(4t+p) - 8\langle d\rangle t + 4t - \langle d\rangle^2 \\ &= \langle d\rangle^2 p + 4t\langle d\rangle^2 - 8t\langle d\rangle + 4t \\ &= \langle d\rangle^2 p + 4t(\langle d\rangle - 1)^2 \approx \langle d\rangle^2 p \ . \end{aligned} \tag{82}$$

If we would repeat the same procedure for the other two estimators, $\hat{\xi}_H(r)$ and $\hat{\xi}_{DP}(r)$, we would get that the bias and variance of these estimators are,

$$\langle 1+\hat{\xi}_{DP}(r)\rangle = \langle d\rangle(1-t) \quad \text{and} \quad Var(\hat{\xi}_{DP}(r)) \approx \langle d\rangle^2(t+p) \quad (83)$$

$$\langle 1+\hat{\xi}_{H}(r)\rangle = \langle d\rangle(1-t) \quad \text{and} \quad Var(\hat{\xi}_{H}(r)) = \langle d\rangle^2 p\ . \quad (84)$$

Now that biases and variances are known for all the estimators, we can compare them, so that we can choose the best estimator for the two-point correlation function. From equations 65 and 66, it can be easily seen that the variance of the Davis&Peebles estimator is proportional to the inverse of the number of data points, $Var(\hat{\xi}_{DP}) \propto 1/N_D$. On the other hand, Hamilton and Landy&Szalay estimators have a variance proportional to the inverse square of the number of data points, $Var(\hat{\xi}_H) = Var(\hat{\xi}_{LS}) \propto 1/N_D^2$. For this reason, we consider that both Hamilton and Landy&Szalay estimators overcome the Davis&Peebles estimator. Furthermore, the Hamilton estimator has a small extra bias introduced by the term (1-t), apart from the one introduced by the integral constraint, compared to the Landy&Szalay estimator and, therefore, we conclude that the best two-point correlation function estimator is the Landy&Szalay one.

Following reference [22], we can confirm our previous conclusion that the Landy&Szalay is the best estimator that can be used to estimate the 2PCF. In this paper, it is also added that when dealing with small scales almost every estimator gives the same result. However, is for large scales where Hamilton and Landy&Szalay estimators outperform the rest. Finally, it is argued that the Hamilton estimator is more sensible to the number of random points, N_R, and thus, the preferred option will finally be the Landy&Szalay estimator.

Now, the projected correlation function is presented, which is a useful tool used to obtain information about the real-space correlation function.

3.3.4 Projected correlation function

From the two-dimensional two-point correlation function, different one-dimensional correlation functions can be inferred. One of them, is the redshift space correlation function, $\xi(s)$, where $s = \sqrt{r_\pi^2 + r_p^2}$ is the redshift space distance between two galaxies. Another important 1D correlation function that can be calculated is the so-called *projected correlation function*, $w_p(r_p)$. This function was first introduced by Davis and Peebles in 1983, in reference

[6], to minimize the effect of peculiar velocities on the correlation functions. They defined the projected correlation function as,

$$w_p(r_p) = \int_{-\infty}^{\infty} \xi^{(z)}(r_\pi, r_p) dr_\pi = \int_{-\infty}^{\infty} \xi(r(r_\pi, r_p)) dr_\pi \ , \tag{85}$$

where $\xi^{(z)}(r_\pi, r_p)$ and $\xi(r(r_\pi, r_p))$ refer to the two-dimensional Cartesian two-point correlation functions in redshift and real space and where, in practice, the integration limits will be changed to a maximum distance along the r_π direction, π_{max}, which will be big enough to include almost every galaxy pair.

As we have assumed the plane-parallel approximation, it can be concluded that the redshift space distortions only affect our targeted galaxies along the r_π direction. Therefore, as the integration is done along the r_π direction no information is lost in the process and, as a consequence, w_p remains unaffected by these distortions. Hence, this projected correlation function is identical to a simple projection of the real-space correlation function, which is given by an Abel transform. Since,

$$r = \sqrt{r_\pi^2 + r_p^2} \qquad \text{and} \qquad dr_\pi = \frac{r dr}{\sqrt{r^2 - r_p^2}} \ , \tag{86}$$

we have that,

$$w_p(r_p) = 2 \int_{r_p}^{\infty} \xi(r) \frac{r dr}{\sqrt{r^2 - r_p^2}} \ . \tag{87}$$

The Abel integral can be inverted to give,

$$\xi(r) = -\frac{1}{\pi} \int_r^{\infty} \frac{w_p'(r_p) dr_p}{\sqrt{r_p^2 - r^2}} \ , \tag{88}$$

where $w_p'(r_p)$ means derivative of $w_p(r_p)$ with respect to r_p. From equation 85 to equation 88, we have followed a completely geometrical procedure and, as no approximations have been done, the result should be, in theory, independent of the chosen scale.

Therefore, we have now found a way to infer the one-dimensional two-point correlation function in real space from the projected correlation function and, hence, from the two-dimensional correlation functions in either

real or redshift space.

In this project, we will derive a code that solves this problem to find whether this process works as it should, for all scales, just for some or if it does not work at all. Henceforth, we will begin showing the results that we have achieved using the Euclid Consortium code in order to find the one-dimensional and two-dimensional correlation functions and, from there, we will move on to derive the deprojected correlation function from the obtained data. After having found the deprojected correlation function, it will be compared to the 1-dimensional two-point correlation function predicted by the Euclid Consortium code.

4 Two point correlation functions

In this section, the results obtained for the different two-point correlation functions will be presented. The section will begin with an introduction to the Minerva simulations. After that, the Euclid Consortium code, which implements the Landy&Szalay estimator, will be used to find the two-point correlation functions in either real or redshift space and one or two dimensions. Then, projected correlation functions will be calculated using the Euclid data previously obtained and, finally, we will proceed to the deprojection of those projected correlation functions.

4.1 Minerva simulations and Euclid code

In section 3.3.3, different estimators were defined to calculate the two-point correlation function. The main conclusion of this section was that the most accurate estimator was the one derived by Landy&Szalay. Hence, two galaxy catalogues have to be simulated if the correlation function wants to be computed, a data catalogue and a random catalogue. The generation of a random catalogue is a crucial part of the 2PCF calculation process. This random set has to represent a Poisson distribution of galaxy points, with the same selection effects as the data catalogue.

This simulation of data and random catalogues was already done, as it can be seen in reference [29], and they have been used within the Euclid Consortium. The results are the so-called Minerva mocks, which are a set of 300 different cosmological mocks produced with N-body simulations. The Euclid satellite will survey the sky from a range of redshifts that goes from 0 to 2. Therefore, Minerva mocks will be stored at five different redshift bins that lie in that range. These mocks have a cubic shape with a $1500h^{-1}$Mpc length in each direction, and periodic boundary conditions are assumed in these boxes.

Minerva mocks contain, approximately, $N_d \approx 4 \times 10^6$ galaxy points and, the number of random points used in the random catalogue will be $N_r = 50N_d$. These numbers take into account a whole Minerva box. However, in practice, we will deal with subdivisions of these boxes due to the long computational time required by a full Minerva box. These subdivisions, called slabs, are created by cutting the cube into nine parallel slices of thickness 1500Mpc/9 $\approx$ 166.7Mpc and thus, the final volume of these slabs will be (1500Mpc) $\times$ (1500Mpc) $\times$ (166.7Mpc). The short dimension of the slab

is oriented along the line-of-sight direction and, as a consequence, r_π will never exceed 166.7Mpc. As all the calculations will be done on these smaller samples, edge effects will have to be taken into account. Furthermore, the number of galaxies in these slabs is approximately $N_d \approx 4.5 \times 10^5$. This number of galaxies is much smaller than the galaxies expected in the Euclid mission, which is about 3×10^7 galaxies, but it does represent a realistic galaxy density.

Once we have the Minerva simulations, we can proceed to the estimation of the two-point correlation function. Euclid Consortium code uses the Landy&Szalay estimator, explained in [46], to calculate different correlation functions or features of these functions such as the multipoles or the projected correlation function. From now on, two-point correlation functions will be estimated using $\Delta r = 1h^{-1}$Mpc bins up to a $r_{max} = 200h^{-1}$Mpc.

Using Minerva simulations, the one-dimensional real-space two-point correlation function can be calculated and, from now on, this correlation function will act as the reference function. This function will be considered as the "correct" one, and it will be compared to the one obtained through the deprojection process to quantify the success or failure of the method. Hence, if the correlation function obtained after the deprojection process equals the one obtained from Minerva simulations, the success of the deprojection method will be proven. This comparison can be made because Minerva catalogues where peculiar velocities are ignored have been developed, which means that these are real-space data sets from where $\xi(r)$ can be calculated directly. However, in real sky surveys, all the measurements belong to redshift space so that the projected correlation function has to be calculated in order to avoid the effect of peculiar velocities and, after that, the deprojection method has to be applied to find the real space $\xi(r)$.

The fact that the real-space correlation function can be predicted from simulated cosmological mocks is a critical point in this project and in the whole Euclid mission. This is mainly due to the fact that the physically meaningful correlation function is the one-dimensional real-space correlation function, $\xi(r)$. From this function, different properties of the large-scale structure of our Universe can be derived and, therefore, it is essential to have a theoretical prediction of this function that can be compared to the one obtained after the deprojection process of real data.

Now, we will move on to apply Euclid Consortium code to calculate all

the correlation functions that are needed in this project.

4.2 Calculating two-point correlation functions using Euclid Code

Making use of the Euclid Consortium code, we are able to calculate different types of correlation functions. We can calculate them in 1 or 2 dimensions and in real or redshift space. Moreover, other properties of the correlation functions can be calculated, such as the different multipoles of these functions, the projected correlation function or the pair counts DD, RR and DR.

We will now move on to show the results of the two-point correlation functions. For all correlation functions, two-hundred different realizations in both real and redshift space have been calculated. Then, the mean overall the realizations was calculated to obtain the most accurate approximation to the true correlation function.

The results begin with the calculation of the real-space two-point correlation function in 1D and for 200 different realizations. In figure 2, it can be seen the 2PCF multiplied by the squared distance scale for only one realization. The multiplication by r^2 is done so that the structure and features of $\xi(r)$ are more easily seen and, from now on, all two-point correlation functions will be plotted using this scale factor. In figure 3 we see the comparison between one realization and the mean over two hundred realizations.

In these plots, several features of the two-point correlation functions can be appreciated. As we can see, the correlation function is quite large for small scales. However, it decays fast and becomes negative for larger scales. Another important feature that is easily visible is the peak around 110Mpc. That peak is provoked by the Baryon Acoustic Oscillations (BAO), which are fluctuations in the visible baryonic matter created by acoustic density waves in the plasma of the early universe. This BAO peak occurs approximately at a comoving separation of $\sim$ 110Mpc as it can be seen from the plots. Moreover, the scale where the BAO peak lies has been proven using CMB physics.

Now, we proceed on the same way to calculate the one-dimensional two-point correlation function in redshift space. In figure 4, we can observe the comparison between one realization of this 2PCF and the mean over 200

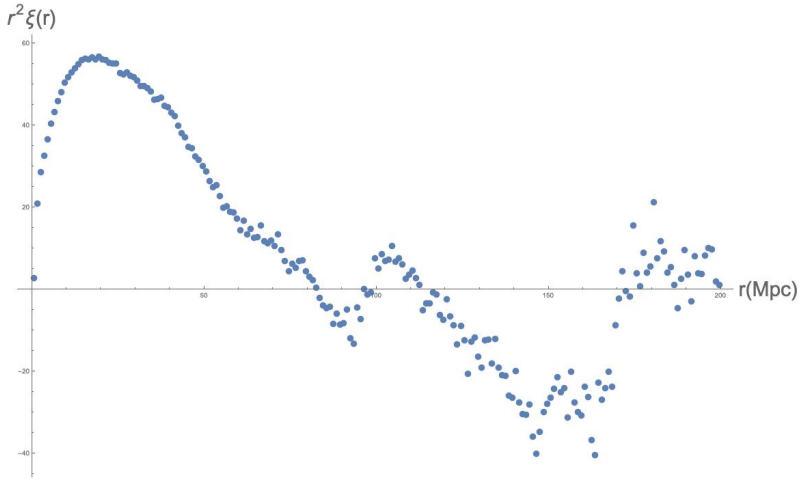

Figure 2: One dimensional two-point correlation function in real space multiplied by the distance scale squared for one realization.

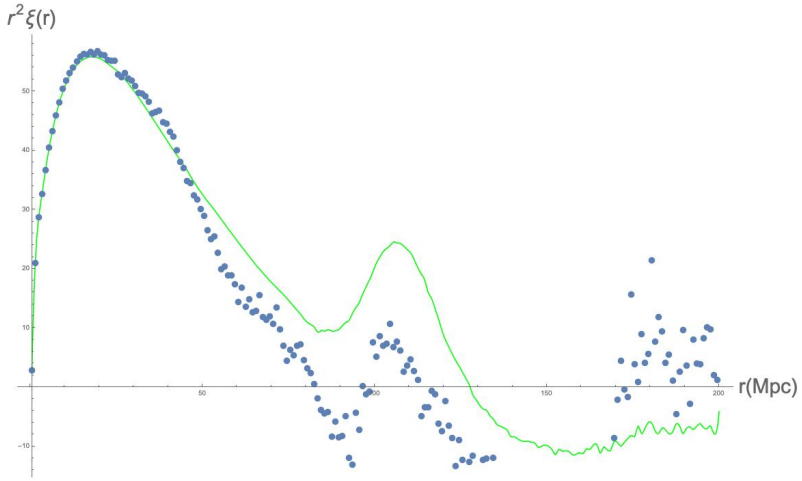

Figure 3: One dimensional two-point correlation function in real space multiplied by the distance scale squared. The green line represents the mean over 200 realizations while the points represent the 2PCF of only one realization.

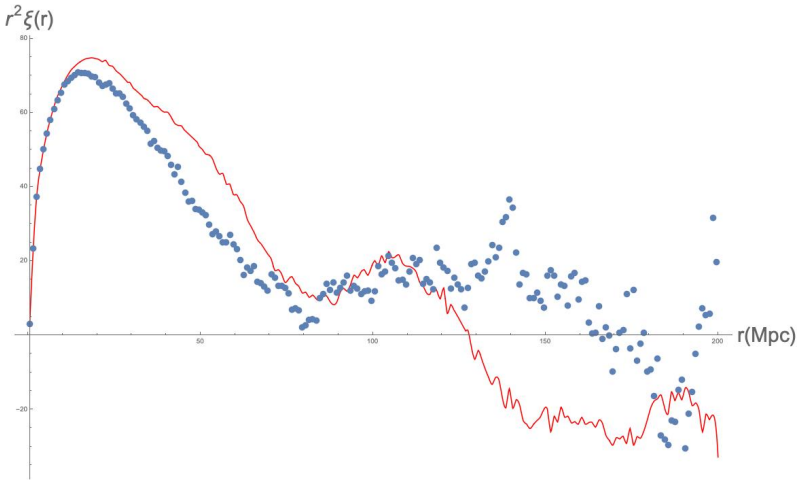

Figure 4: One dimensional two-point correlation function in redshift space multiplied by the distance scale squared. The red line represents the mean over 200 realizations while the points represent the 2PCF of only one realization.

realizations. At first sight, we could say that the shape of this correlation function is practically the same as the one in real space.

We now compare these two functions to observe their differences. This can be seen in figure 5. The main difference between these correlation functions comes in small scales. Here, the redshift space correlation function is slightly bigger than the one in real space. This effect is mainly produced by the peculiar velocities of individual galaxies that distort the true pattern of clustering. For larger scales, it can also be seen that the correlation function in redshift space is more negative than the one in real space. However, this effect is mainly produced by the scaling, r^2, we have chosen to represent our correlation functions. We can relate the two-point correlation function in redshift space to that function in real space in the region where linear perturbation theory can be applied and treating the correlation functions as if they would follow a pure power-law shape. As it is stated in references [10] and [31], this relation is given by,

$$\frac{\xi(s)}{\xi(r)} = 1 + \frac{2}{3}\beta + \frac{1}{5}\beta \ , \tag{89}$$

where s is the distance in redshift space and beta is a factor that takes

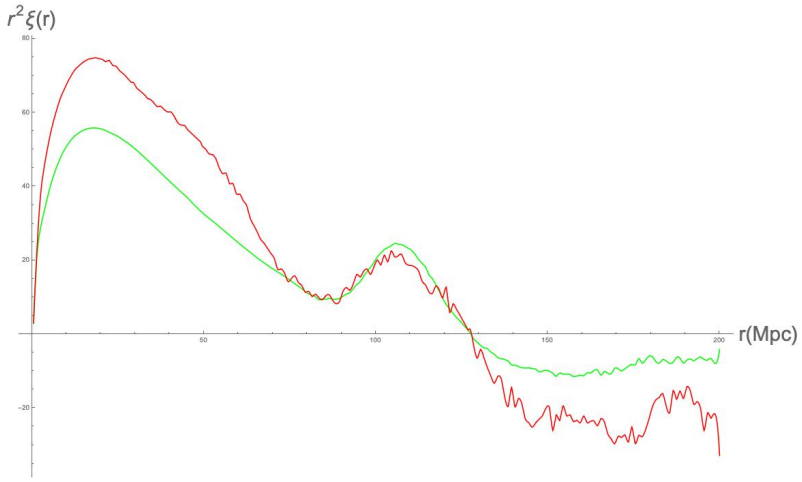

Figure 5: One dimensional 2PCF in real and redshift space multiplied by the distance scale squared. The red line corresponds to the 2PCF in redshift space while the green one corresponds to real space.

into account the large-scale effects of linear redshift-space distortions and is defined by $\beta = \Omega^{0.6}/b$, with Ω being the cosmological density parameter and b the bias parameter that relates fluctuations in mass with those in the galaxy distribution $\left(\frac{\delta\rho}{\rho}\right)_g = b\left(\frac{\delta\rho}{\rho}\right)_m$.

Now that we have found one-dimensional correlation functions in both spaces, we will move on to calculate the same results but for two dimensions. Again, we have calculated two hundred realizations of both real and redshift space correlation functions. This has taken a long computational time to run. For a single realization in one dimension, the code takes, in general, almost 310 seconds to run, while for two dimensions the code takes approximately 860 seconds per realization.

After calculating all the 2PCF for each realization, we obtained the following results. In real space, for only one realization, the result can be seen in figure 6. Moreover, in figure 7 we can see the 2PCF in two-dimensional real space for the mean of two hundred realizations. The first feature that can be seen in this last figure is the fact that we have circular contours of equal ξ. This is explained by the fact that, by isotropy, the correlation function ξ should only depend on the scalar separation between galaxies. These contours indicate the constant probability of finding galaxy pairs at a given

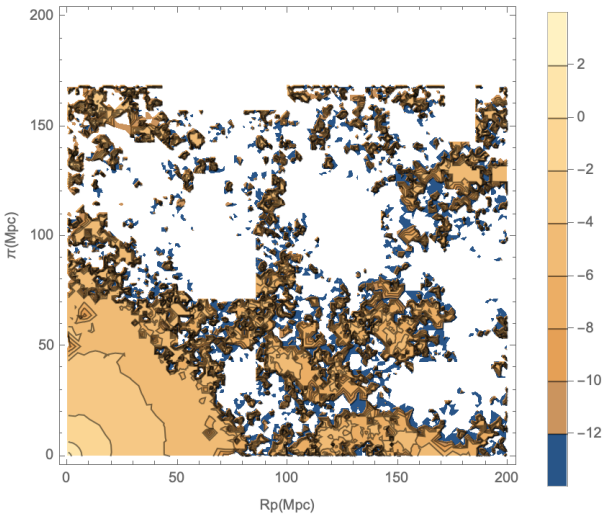

Figure 6: Two-dimensional two-point correlation function in real space for 1 realization. The colour box show the value of the $\log[\xi(r_p, \pi)]$ so that the differences can be easily seen. White colour indicates negative correlation, except for the upper rectangular part of the plot, where we have no data.

distance r_p and π. From now on, both r_π and π represent the parallel to the line-of-sight direction. Another feature that we can see in figure 7, is the fact that for high values of π and small values of r_p, we have some random scatter objects, due to the limited number of data pairs, that differ from the circular contours explained before. These random scatter points could be eliminated by calculating more realizations of the 2D 2PCF and calculating the mean over all of them. The last important aspect of these plots that is easily visible, is the white rectangle on the upper part of the plots. While in these plots white colour means negative correlations, in these rectangles the meaning of white colour is different. In these rectangles, we do not have any more data points and therefore, we assume the correlation to be exactly 0. The lack of data points in this area is strictly related to the slab geometry of our data sets. As it was explained before, we have been using slices of the full Minerva box called slabs, and these slabs have a maximum distance along the line-of-sight direction of $\pi \approx 166.7$Mpc. Hence, there will be no data once that distance is exceeded.

In the same way we did for one-dimensional correlations, now we plot the two-dimensional correlation function in redshift space. We will plot directly

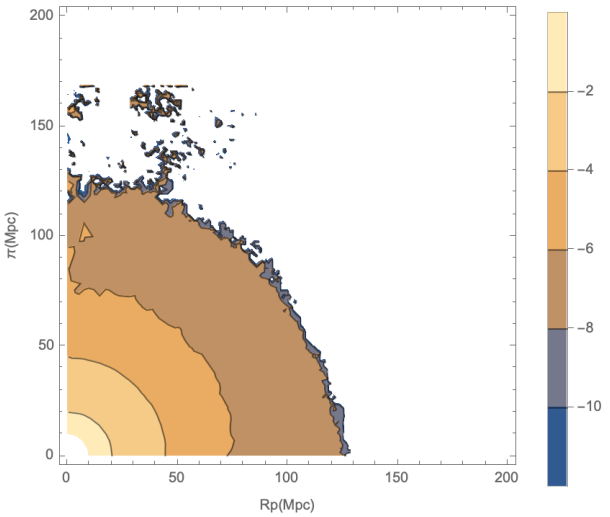

Figure 7: Two-dimensional two-point correlation function in real space. The colour box show the value of the $\log[\xi(r_p, \pi)]$ so that the differences can be easily seen. White colour indicates negative correlation, except for the upper rectangular part of the plot, where we have no data.

the mean over two hundred different realizations. The result can be seen in figure 8. In this figure, again, there are some features worth mentioning. As we had in real space two-dimensional correlation functions, we have an upper white rectangle, which means the lack of data-data, DD, points. Also, it can be seen how peculiar velocities distort the two-dimensional two-point correlation function in redshift space in small scales. In this case, peculiar velocities distort the contours elongating them along the line of sight direction. It exhibits another distortion effect caused by the velocity of dispersion in collapsed objects, which is called *finger of God* and is a radially elongated structure created by non-linear effects that appears to be pointing towards the observer. This effect is easily visible in the Coma cluster as it is exposed in figure (1) of reference [16]. Furthermore, we can also see how at large scales, the correlation function is compressed along the parallel to the line of sight direction, π, due to the coherent large-scale streaming as galaxies fall into the potential wells.

We have already found all the two-point correlation functions that are needed in this project. We have two-point correlation functions in either real or redshift space and in 1 or 2 dimensions. Using these data, we will

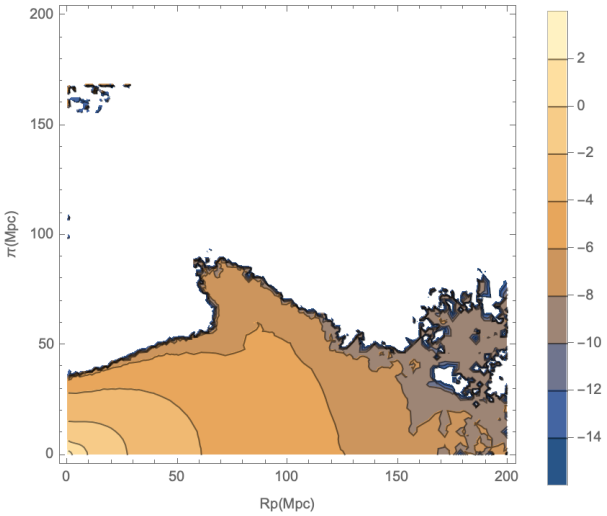

Figure 8: Two-dimensional two-point correlation function in redshift space. The colour box show the value of the $\log[\xi(r_p, \pi)]$ so that the differences can be easily seen. White colour indicates negative correlation, except for the upper rectangular part of the plot, where we have no data.

now move on to calculate projected correlation functions and then, we will proceed to the deprojection method. From here, we will be able to demonstrate if we can get the 2PCF in one-dimensional real space from the two-dimensional 2PCF in redshift space, which is the data we will get from the Euclid satellite.

4.3 Calculating the projected correlation function $w_p(r_p)$

As it was explained in section 3.3.4, the projected correlation function is computed to deal with the errors in the measured signal produced by redshifts. When the anisotropic two-point correlation function $\xi(r_p, \pi)$ was computed, it was found that the r_p direction was almost unaffected by redshift distortions while, on the other hand, the direction along the line-of-sight was distorted. Therefore, in order to avoid these distortions an integration has to be carried out along the π direction, as it is stated in equation 85. As a result, the projected correlation function will be obtained and we will have eliminated the redshift-space distortions without losing any valuable information.

In order to calculate the projected correlation function, an important choice has to be made, the maximum integration distance π_{max}. This is not an easy question to solve because if we introduce a low π_{max}, we will not be counting some galaxy pairs and therefore, the final result will not be accurate enough. On the other hand, if we put a very high integration limit, the final result will be distorted. As a conclusion, we will calculate different projected correlation functions for different values of π_{max}. Then, as we have already predicted the one-dimensional two-point correlation function in real space using Euclid code, we will make use of this data and equation 87 to calculate the "true" projected correlation function. Then, we will compare it to the ones calculated through the integration of real-space $\xi(r_p, \pi)$ so that we can fix the maximum integration limit. Hence, making use of the 2D 2PCF obtained through the Euclid code and the definition of the projected correlation function, in equation 85, we have calculated $w_p(r_p)$ using 100, 125, 150 and 175 Mpc as the maximum integration limit. It has no sense to integrate till distances higher than 175Mpc because this distance already exceeds the maximum distance along the line-of-sight direction allowed by our slabs, which was $\pi \approx 166.7$Mpc. It is important to notice that from Euclid code we obtain a *.txt* data file with three data columns r_p, π and $\xi(r_p, \pi)$. Therefore, we are given the value of the anisotropic correlation function for different specific distances and thus, the projected correlation function will only be calculated for specific distances within the range of variation of r_p, which means that we will not have a continuous function for w_p. In figure 9, the results for these π_{max} and the $w_p(r_p)$ obtained through equation 87 can be seen. In the plot, the continuous line is the "correct" $w_p(r_p)$ while the dots belong to different integrations of real-space $\xi(r_p, \pi)$. As it can be seen in the plot, there are no big differences when $\xi(r_p, \pi)$ is integrated using different π_{max}. From the theory of projected correlated functions, we should obtain the same projected correlation function when integrating $\xi^{(z)}(r_p, \pi)$ or $\xi(r_p, \pi)$. However, as it can be seen in figure 10, when we integrate the anisotropic correlation function in redshift space, the results vary much more than in the real space integration.

From these two plots, it can be seen that when $\xi^{(z)}(r_p, \pi)$ is integrated along the π direction, the result is a function which has a bigger magnitude for big scales than the one obtained when $\xi(r_p, \pi)$ is integrated along the same direction. From these resuts, a precise π_{max} can not be concluded. However, a good guess would be a $\pi_{max} \approx 150 - 175$ Mpc, because a lower π_{max} would give as a result a w_p too different to the one obtained from the integration of $\xi(r)$ in equation 87. Later on, it will be shown that when large

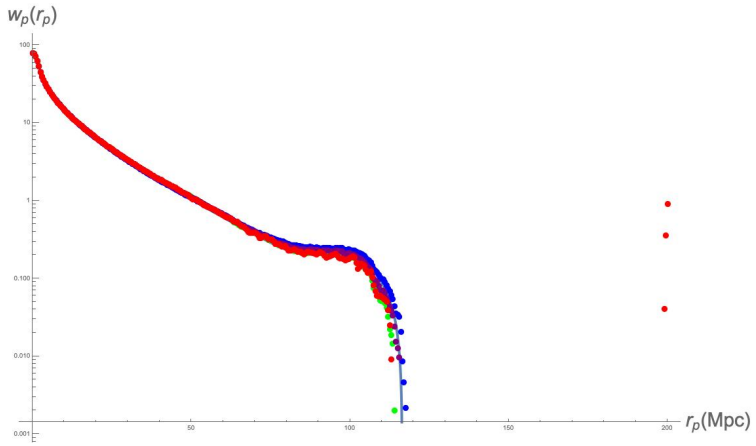

Figure 9: Different projected correlation functions obtained by integrating $\xi(r_p, \pi)$. Blue dotted line refers to a w_p calculated with $\pi_{max} = 100$Mpc, purple dots to $\pi_{max} = 125$Mpc, green dots to $\pi_{max} = 150$Mpc and red dots to $\pi_{max} = 175$Mpc. The continuous line represents the "correct" w_p.

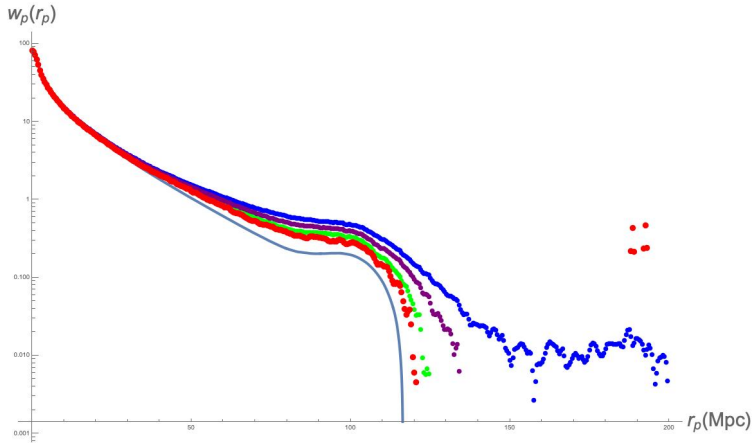

Figure 10: Different projected correlation functions obtained by integrating $\xi^{(z)}(r_p, \pi)$. Blue dotted line refers to a w_p calculated with $\pi_{max} = 100$Mpc, purple dots to $\pi_{max} = 125$Mpc, green dots to $\pi_{max} = 150$Mpc and red dots to $\pi_{max} = 175$Mpc. The continuous line represents the "correct" w_p.

values of π_{max} are used, after the deprojection process, a very distorted $\xi(r)$ is obtained. On the other hand, if a very low π_{max} is used, a precise but not accurate result will be obtained, where precision is defined as the closeness of the measurements to each other and accuracy refers to the closeness of the measurements to the true or real value.

Furthermore, it could also be concluded that the possible errors that we may find in the final deprojected two-point correlation functions are not or not only due to the deprojection process itself but to the difference on the projected correlation functions. Comparing the results obtained for the different projected correlation functions, it can be seen that up to a distance of around 40Mpc, the results are the same when we integrate the real-space or redshift-space anisotropic correlation function. After that distance, both results begin to differ for an unknown reason. One of the possible explanations for these differences could be the slab geometry of our simulations. However, in order to prove this statement, we would need to repeat all the process using the full-box geometry instead of the slab geometry. This would eliminate the edge effects on our correlation functions and could, in theory, minimise these differences. However, it would take enormous computational time. Each computation of the anisotropic correlation function using a full-box geometry takes approximately seven hours to be computed which makes a proper calculation using these correlation functions impossible. If we would have wanted to compute 200 different realizations of the anisotropic correlation function for the whole Minerva box, around 58 days of computation would have been needed, and this would have been only in real or redshift space. Therefore, we will continue using the results obtained through the slab geometry which, even though they are not completely accurate, are good enough results taking into account the computational time they require.

In this subsection, different projected correlation functions have been calculated for different values of the maximum integration distance π_{max}. The result has not been conclusive, and a proper π_{max} has not been found yet. Hence, the deprojection procedure will be carried out for all the projected correlation functions. After that, the results will be compared to the $\xi(r)$ obtained from Euclid code.

4.4 Deprojection procedure

Finally, we have arrived at the final step of the deprojection procedure. Till now, we have found the anisotropic two-point correlation functions in real or redshift space and from them, we have calculated the projected correlation functions. From the Minerva simulations a theoretical one-dimensional two-point correlation function was found and, comparing the theoretical results with the projected correlation functions a proper value for the maximum integration limit was not found. Therefore, the deprojection procedure will now be applied using different projected correlation functions found in the previous section.

The deprojection formula, shown in 88, makes use of the derivative of the projected correlation function. In the code that has been developed for this project, because the projected correlation function has been calculated for different distances in the interval $[0, 200]$Mpc with a separation of $h = 1Mpc$, a simple centre divided difference method has been used to calculate the first derivative. This numerical method calculates the derivative by,

$$w_p'(x) = \frac{w_p(x+h) - w_p(x-h)}{2h} \ , \tag{90}$$

and the results can be seen in figure 11. In this plot, it can be seen that the derivatives of the projected correlation function remain practically the same for every previous choice of integration limit π_{max}. However, it can also be appreciated that when the integration limit is increased, some noise is introduced in the results making them have a more oscillatory behaviour, especially for large scales, which makes the results less precise. In this case, we have calculated the derivatives of the projected correlation functions obtained integrating the real-space anisotropic correlation function, but if we would have used the ones obtained from the integration of the redshift-space anisotropic correlation function the results would have had a similar behaviour.

Once the first derivative is calculated we have to compute the Abel integral to finish the deprojection process. In this integral, we have a factor $1/\sqrt{y^2 - r^2}$ so that, when the deprojected correlation function is computed, we will be integrating a function that diverges at one end-point at $y = r$. Therefore, a very significant contribution to the integral comes from the region close to this end-point $y = r$. However, if the integral is computed as a sum over discrete bins, which is the method we used to compute the

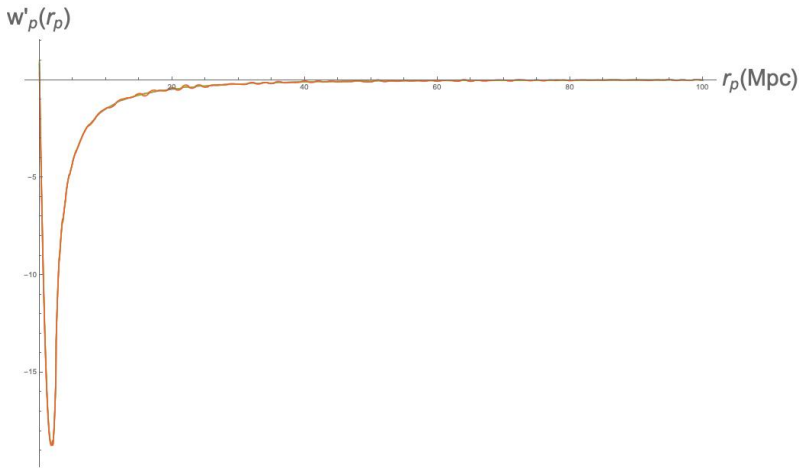

Figure 11: First derivative of the projected correlation function w_p calculated with the same code of colours previously used. Blue line for $\pi_{max} = 100$Mpc, purple line for $\pi_{max} = 125$Mpc, green line for $\pi_{max} = 150$Mpc and red line to $\pi_{max} = 175$Mpc. Interpolation has been used between the points where the projected correlation function is known.

projected correlation function, this important contribution would be lost reaching to a wrong result.

In order to solve this problem with the integration, we have divided the integration range in much smaller bins. The new bin size will be 0.01Mpc so that we have 100 sub-bins per Mpc. Evaluating the factor $1/\sqrt{y^2 - r^2}$ in these sub-bins is not a problem but, evaluating the derivative of the projected correlation function is not that obvious. The solution came by evaluating the derivative in the middle points of the original data grid, following equation 90, and then interpolating between those points. Spline interpolation method has been used to evaluate the derivative on the new grid. When the interpolation is finished we have as a result a new grid with a bin size of 0.01Mpc instead of the previous 1Mpc. Finally, the integral is computed as a sum over these new discrete sub-bins.

The results for the deprojection procedure are now presented. In figure 12, we can see the deprojected correlation functions obtained from the real-space anisotropic correlation function, $\xi(r_p, \pi)$, and the one obtained directly from Minerva simulations using Euclid code. From the plot, it is

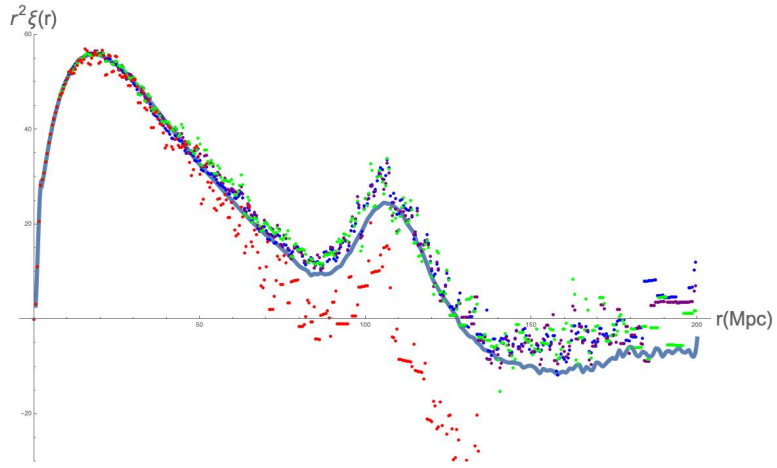

Figure 12: Dotted lines represent the deprojected correlation functions obtained from the real-space anisotropic correlation function. As before, blue dots correspond to a choice of π_{max} = 100Mpc, purple dots to π_{max} = 125Mpc, green dots to π_{max} = 150Mpc and red dots to π_{max} = 175Mpc. The continuous blue line represents the "correct" $\xi(r)$ obtained from the Euclid Minerva simulations.

obvious that the deprojection procedure works correctly for small scales, approximately until 40Mpc. However, for larger scales, some noise is introduced in the results, and this noise increases with the integration limit used in the first step of the deprojection procedure π_{max}. Hence, we can conclude that π_{max} = 175Mpc is not a good choice as an integration limit due to the noise that introduces in our results at large scales. Apart from that, the results for the other choices of π_{max} 100Mpc, 125Mpc and 150Mpc, show no bigger differences among them, which do not help us identify a proper maximum integration limit.

In figure 13, the deprojected correlation functions obtained from the redshift-space anisotropic correlation function, $\xi^{(z)}(r_p, \pi)$, can be seen compared to the "correct" one-dimensional two-point correlation function obtained from the Minerva simulations. In this plot, in contrast to figure 12, higher precision is achieved. However, the accurateness of the results is lower than in the previous plot. Again, for small scales the behaviour of the deprojected correlation functions is correct until 40Mpc but, for larger scales, does not provide good results. It is especially visible how in this case, in the

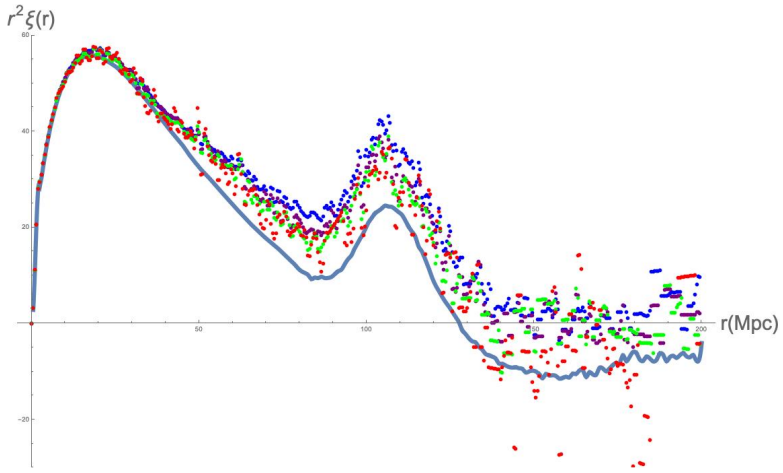

Figure 13: Dotted lines represent the deprojected correlation functions obtained from the redshift-space anisotropic correlation function. As before, blue dots correspond to a choice of π_{max} = 100Mpc, purple dots to π_{max} = 125Mpc, green dots to π_{max} = 150Mpc and red dots to π_{max} = 175Mpc. The continuous blue line represents the "correct" $\xi(r)$ obtained from the Euclid Minerva simulations.

deprojected correlation functions, the BAO peak is more pronounced than in the one-dimensional real-space correlation function $\xi(r)$obtained from the Minerva simulations. When the deprojection was done using $\xi(r_p, \pi)$ as the initial data, figure 12, this peak resembled better, though not perfect, to the one predicted by Minerva simulations. Therefore, from this fact, we can conclude that the results increase their accuracy when we begin with the real-space anisotropic correlation function $\xi(r_p, \pi)$.

Furthermore, from figure 12 we can obtain more information than the fact that π_{max} = 175Mpc is a too large integration limit. Comparing the other three deprojected correlation functions, the ones obtained with $\pi_{max} = 100$Mpc, $\pi_{max} = 125$Mpc and $\pi_{max} = 150$Mpc, to the one obtained directly from the Minerva simulations and Euclid code, we can see that the deprojected correlation function that behaves more closely to the "correct" one is the one obtained with a maximum integration limit $\pi_{max} = 150$Mpc. This is a logical conclusion because integrating till 150Mpc, we will count more galaxy pairs than in the case of 100Mpc or 125Mpc, so the results should, in theory, be more accurate. Furthermore, integrating till 150 Mpc

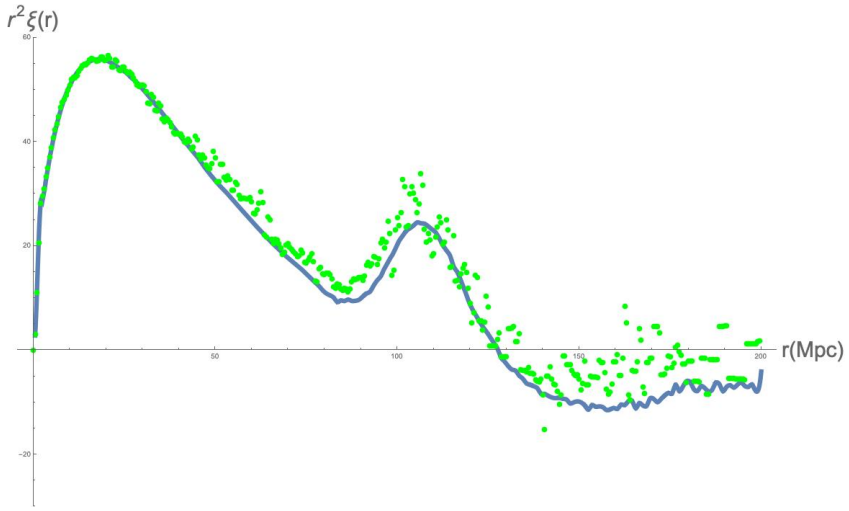

Figure 14: Green dots represent the deprojected correlation function obtained from the real-space anisotropic correlation function with a $\pi_{max} =$ 150Mpc, while the continuous blue line represents the "correct" $\xi(r)$ obtained from the Euclid Minerva simulations.

does not introduce much noise in the results and thus, the deprojected correlation function has a good enough precision, making it behave correctly. Henceforth, we will consider that the best choice for the integration limit in equation 85, is $\pi_{max} = 150$Mpc because it gives the closest results to the ones shown as a green line in figure 3 and, at the same time, does not introduce much noise in the results as it happens when the integration limit is increased. In figure 14 it can be seen the comparison between the deprojected correlation function obtained with $\pi_{max} = 150$Mpc and the one obtained directly from the Euclid code. In that plot, it can be seen how the results of the deprojection procedure are close to the real values of the real-space two-point correlation function.

On the other hand, if we take the deprojected correlation function obtained from the redshift-space anisotropic correlation function $\xi^{(z)}(r_p, \pi)$ with $\pi_{max} = 150$Mpc, it can be seen that the results differ for large scales where the deprojected correlation function presents higher values than those of the "correct" two-point correlation function calculated directly from Minerva simulations. These results can be seen in figure 15.

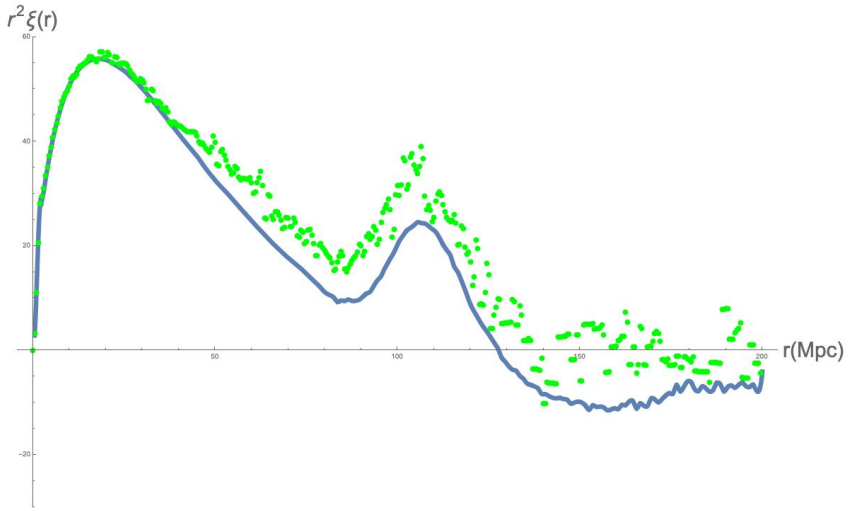

Figure 15: Green dots represent the deprojected correlation function obtained from the redshift-space anisotropic correlation function with a $\pi_{max} =$ 150Mpc, while the continuous blue line represents the "correct" $\xi(r)$ obtained from the Euclid Minerva simulations.

Taking everything into account, we can conclude that the deprojection procedure works fine for small scales up to 40Mpc not matter whether we have begun from the real-space or redshift-space anisotropic correlation function. From that distance scale onwards, it does matter whether we begin with real-space or redshift-space anisotropic correlation function. If the initial data is a two-dimensional two-point correlation function in real space the final result of the deprojection procedure will be very close to the true value of the one-dimensional two-point correlation function in real-space. However, if the initial data is a redshift-space anisotropic correlation function, the deprojected correlation function will differ from the one we have assumed as correct for large scales. It can also be concluded that these differences are not provoked by the deprojection procedure itself, because we had previously found that the projected correlation functions obtained from real-space or redshift-space anisotropic correlation functions were not equal, as it can be seen in figures 9 and 10, as they should have been.

5 Conclusions

Throughout this project, a topic that is really under discussion within the scientific community nowadays has been discussed.

We began this work by introducing the theory behind a homogeneous Universe. Friedmann equations, Hubble parameter, Hubble expansion and redshift concepts were introduced and, all these ideas helped us when the inhomogeneous Universe was introduced. The concept of inhomogeneity was explained in terms of the density perturbation field $\delta(\mathbf{x})$ and from this density perturbation field, and with the linear perturbation theory, the equations for the motion of a perfect fluid were derived.

After that, all the theory behind the two-point correlation functions was introduced. In this part of the theoretical introduction, the most important concepts needed in this project were introduced including the deprojection procedure. The validation with real simulated data of this deprojection procedure represents the main objective of this project.

The practical part of the project began calculating different realizations of the one-dimensional or two-dimensional two-point correlation function in either real-space or redshift-space. This 2PCF were calculated using Minerva simulations, which are a set of three hundred cosmological mocks, as the input data and the code developed by the Euclid consortium. We calculated two hundred realizations of each two-point correlation function $\xi(r)$, $\xi^{(z)}(s)$, $\xi(r_p, \pi)$ and $\xi^{(z)}(r_p, \pi)$ and took the mean over them to obtain the most realistic functions.

From one-dimensional correlation functions, we observed that the redshift space 2PCF, for small scales, took higher values than the one in real space and, we concluded that this effect was due to the peculiar velocities of the galaxies. The same procedure was followed for anisotropic correlation functions and, in this case, we found that for $\xi(r_p, \pi)$ we obtained circular contours of equal ξ, which was explained by assuming an isotropic Universe so that function ξ should only depend on the scalar separation between galaxies. Therefore, the circular contours we obtained represent the constant probability of finding galaxy pairs at a given distance r_p and π. On the other hand, a new distorted pattern was obtained for the anisotropic correlation function in redshift-space. In this case, two different kinds of distortions were found for different scales. For small scales, peculiar veloci-

ties distorted the contours elongating them along the line-of-sight direction. However, at large scales, it could be seen a compression along the π direction that was due to the coherent large-scale streaming of galaxies falling into the potential wells.

Once all functions were calculated, the deprojection process was begun. The aim of this procedure was quite simple, obtaining real-space two-point correlation functions from the data obtained from sky surveys, which is redshift-space data. The real-space correlation function $\xi(r)$ is the function that is physically meaningful and, therefore, a precise method was needed to calculate this function. The deprojection process began by calculating the projected correlation function. This function eliminates the distortions introduced by peculiar velocities by integrating along the line-of-sight direction both the real-space of redshift-space anisotropic correlation functions. During the integration, a maximum integration limit π_{max}, which was later fixed to 150Mpc, had to be chosen. At the beginning of the procedure, we had no previous information about the proper integration limit and thus, different projected correlation functions were found by changing π_{max}. From the theory of projected correlation functions, we knew that the projected correlation functions obtained when real-space or redshift-space anisotropic correlation functions were integrated till the same π_{max} had to be equal. However, the results obtained in this project proved it wrong. The results showed different $w_p(r_p)$ and the most probable explanation for these differences was due to the fact that at the moment of the calculation of the two-point correlation functions, a slab geometry was used in our simulations. This meant that the correlation functions were not calculated for the whole Minerva box but for a smaller part of it called slab. If we would have wanted to correct these differences among the projected correlation functions, we would have needed to calculate every realization for the whole Minerva box which would have implied a much longer computational time that would have made this project impossible.

After the calculation of the different projected correlation functions, the last step of the deprojection procedure was applied. We calculated the deprojected correlation functions for each projected correlation function we had. As it was expectable, the results again differed from one another. The deprojected correlation functions obtained from the integration of $\xi(r_p, \pi)$ showed a lower precision than the ones obtained from $\xi^{(z)}(r_p, \pi)$ but, at the same time, their accuracy was higher.

One aspect that deserves comment is the fact that till 40Mpc, both the deprojected correlation functions and the real-space correlation function calculated directly with the Euclid code matched almost perfectly, from where it was assumed that the deprojection procedure worked fine at least till those distances. When the distance scale was increased, differences began to appear but we could assume that those differences between the expected two-point correlation function and the ones obtained through the deprojection process were not due to a fault on the deprojection procedure, because the projected correlation functions used on the procedure were already different when they should have not been.

Taking everything into account, during this project we have found that the deprojection procedure presents no problems till distances up to 40Mpc. After that distance, a difference between the expected two-point correlation function and the deprojected correlation function arises. This difference can be due to a fault in the deprojection procedure but, due to its geometrical nature and the fact that no approximations are done in the process, the most probable explanation for these differences is given by the edge effects introduced by the slab geometry of the Minerva simulations used to calculate the prior anisotropic two-point correlation functions.

As a final note, Euclid satellite will measure the large-scale structure of the Universe out to redshifts of 2, which means that the measurements will cover the large-scale structure evolution from nowadays till the period when the Universe was one-third of its current age, with the highest precision ever achieved. These measurements will cover the period where dark energy accelerated the expansion of the Universe and thus, it will enable us to answer lots of questions that are still unsolved in these days. In order to measure the large-scale structure of the Universe, two-point correlation functions will be used. Hence, the importance of understanding the behaviour of two-point correlation functions is capital for this space mission as well as having a precise method for obtaining real-space correlation functions from redshift-space data.

6 Bibliography

References

[1] S. Weinberg, "Gravitation and cosmology: principles and applications of the general theory of relativity" (Wiley, New York,1972).

[2] H. Mo, F. van den Bosch, and S. White: Galaxy Formation and Evolution (Cambridge University Press 2010)

[3] S. Dodelson: Modern Cosmology (Academic Press 2003)

[4] S. D. Landy and A. S. Szalay, Astrophys. J. **412** (1993) 64.

[5] A. L. Coil, arXiv:1202.6633 [astro-ph.CO].

[6] M. Davis and P. J. E. Peebles, Astrophys. J. **267** (1982) 465.

[7] A. J. S. Hamilton, Astrophys. J. **417** (1993) 19. doi:10.1086/173288

[8] K. B. Fisher, M. Davis, M. A. Strauss, A. Yahil and J. P. Huchra, Mon. Not. Roy. Astron. Soc. **267** (1994) 927

[9] E. Keihänen *et al.*, arXiv:1905.01133 [astro-ph.CO].

[10] N. Kaiser, Mon. Not. Roy. Astron. Soc. **227** (1987) 1.

[11] I. Zehavi *et al.* [SDSS Collaboration], Astrophys. J. **571** (2002) 172

[12] A. Veropalumbo, F. Marulli, L. Moscardini, M. Moresco and A. Cimatti, Mon. Not. Roy. Astron. Soc. **458** (2016) no.2, 1909

[13] F. Marulli, D. Bianchi, E. Branchini, L. Guzzo, L. Moscardini and R. E. Angulo, Mon. Not. Roy. Astron. Soc. **426** (2012) 2566

[14] D. W. Hogg, astro-ph/9905116.

[15] D. F. Watson, A. A. Berlind and A. R. Zentner, Astrophys. J. **738** (2011) 22 doi:10.1088/0004-637X/738/1/22 [arXiv:1101.5155 [astro-ph.CO]].

[16] V. de Lapparent, M. J. Geller and J. P. Huchra, Astrophys. J. **302** (1986) L1. doi:10.1086/184625

[17] ttps://www.astro.rug.nl/ weygaert/tim1publication/lss2009/lss2009.linperturb.pdf

[18] L. Salas and I. Cruz-Gonzlez, Rev. Mex. Astron. Astrofis. **55** (2019) no.1, 93 [arXiv:1810.01803 [astro-ph.CO]].

[19] C. M. Baugh, Mon. Not. Roy. Astron. Soc. **280** (1996) 267 doi:10.1093/mnras/280.1.267 [astro-ph/9512011].

[20] M. Vargas-Magana *et al.*, Astron. Astrophys. **554** (2013) A131 doi:10.1051/0004-6361/201220790 [arXiv:1211.6211 [astro-ph.CO]].

[21] D. Jeong, L. Dai, M. Kamionkowski and A. S. Szalay, Mon. Not. Roy. Astron. Soc. **449** (2015) no.3, 3312 doi:10.1093/mnras/stv525 [arXiv:1408.4648 [astro-ph.CO]].

[22] M. Kerscher, I. Szapudi and A. S. Szalay, Astrophys. J. **535** (2000) L13 doi:10.1086/312702 [astro-ph/9912088].

[23] A. J. S. Hamilton and M. Culhane, Mon. Not. Roy. Astron. Soc. **278** (1996) 73 doi:10.1093/mnras/278.1.73 [astro-ph/9507021].

[24] Y. S. Loh *et al.*, Mon. Not. Roy. Astron. Soc. **407** (2010) 55 doi:10.1111/j.1365-2966.2010.16908.x [arXiv:1004.3382 [astro-ph.CO]].

[25] E. Castorina and M. White, Mon. Not. Roy. Astron. Soc. **476** (2018) no.4, 4403 doi:10.1093/mnras/sty410 [arXiv:1709.09730 [astro-ph.CO]].

[26] D. Jeong, L. Dai, M. Kamionkowski and A. S. Szalay, Mon. Not. Roy. Astron. Soc. **449** (2015) no.3, 3312 doi:10.1093/mnras/stv525 [arXiv:1408.4648 [astro-ph.CO]].

[27] A. Cimatti *et al.*, arXiv:0912.0914 [astro-ph.CO].

[28] J. N. Grieb *et al.* [BOSS Collaboration], Mon. Not. Roy. Astron. Soc. **467** (2017) no.2, 2085 doi:10.1093/mnras/stw3384 [arXiv:1607.03143 [astro-ph.CO]].

[29] J. N. Grieb, A. G. Snchez, S. Salazar-Albornoz and C. Dalla Vecchia, Mon. Not. Roy. Astron. Soc. **457** (2016) no.2, 1577 doi:10.1093/mnras/stw065 [arXiv:1509.04293 [astro-ph.CO]].

[30] Y. Zu, Z. Zheng, G. T. Zhu and Y. P. Jing, Astrophys. J. **686** (2008) 41 doi:10.1086/591071 [arXiv:0712.3570 [astro-ph]].

[31] N. P. Ross *et al.*, Mon. Not. Roy. Astron. Soc. **381** (2007) 573 doi:10.1111/j.1365-2966.2007.12289.x [astro-ph/0612400].

[32] J. da Angela, P. J. Outram, T. Shanks, B. J. Boyle, S. M. Croom, N. S. Loaring, L. Miller and R. J. Smith, Mon. Not. Roy. Astron. Soc. **360** (2005) 1040 doi:10.1111/j.1365-2966.2005.09094.x [astro-ph/0504438].

[33] J. Carrick, S. J. Turnbull, G. Lavaux and M. J. Hudson, Mon. Not. Roy. Astron. Soc. **450** (2015) no.1, 317 doi:10.1093/mnras/stv547 [arXiv:1504.04627 [astro-ph.CO]].

[34] Hannu Kurki-Suonio, Galaxy Survey Cosmology, Lecture notes, 2019. http://www.courses.physics.helsinki.fi/teor/gsc/ .